



Methods for evaluating information sources

An annotated catalogue

Hjørland, Birger

Published in:
Journal of Information Science

Publication date:
2012

Document version
Early version, also known as pre-print

Citation for published version (APA):
Hjørland, B. (2012). Methods for evaluating information sources: An annotated catalogue. *Journal of Information Science*, 38(3), 258-268. <http://jis.sagepub.com/content/38/3/258.full.pdf+html>

Journal of Information Science

<http://jis.sagepub.com/>

Methods for evaluating information sources: An annotated catalog

Birger Hjørland

Journal of Information Science published online 18 April 2012

DOI: 10.1177/0165551512439178

The online version of this article can be found at:

<http://jis.sagepub.com/content/early/2012/04/16/0165551512439178>

Published by:



<http://www.sagepublications.com>

On behalf of:



[Chartered Institute of Library and Information Professionals](#)

Additional services and information for *Journal of Information Science* can be found at:

Email Alerts: <http://jis.sagepub.com/cgi/alerts>

Subscriptions: <http://jis.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [OnlineFirst Version of Record](#) - Apr 18, 2012

[What is This?](#)

Methods for evaluating information sources: An annotated catalogue

Birger Hjørland

Royal School of Library and Information Science, Denmark

Journal of Information Science

1–11

© The Author(s) 2012

Reprints and permission: sagepub.

co.uk/journalsPermissions.nav

DOI: 10.1177/0165551512439178

jis.sagepub.com



Abstract

The article briefly presents and discusses 12 different approaches to the evaluation of information sources (for example a Wikipedia entry or a journal article): (1) the checklist approach; (2) classical peer review; (3) modified peer review; (4) evaluation based on examining the coverage of controversial views; (5) evidence-based evaluation; (6) comparative studies; (7) author credentials; (8) publisher reputation; (9) journal impact factor; (10) sponsoring: tracing the influence of economic, political, and ideological interests; (11) book reviews and book reviewing; and (12) broader criteria. Reading a text is often not a simple process. All the methods discussed here are steps on the way on learning how to read, understand, and criticize texts. According to hermeneutics it involves the subjectivity of the reader, and that subjectivity is influenced, more or less, by different theoretical perspectives. Good, scholarly reading is to be aware of different perspectives, and to situate oneself among them.

Keyword

critical research assessment; evaluation of information sources; source criticism

1. Introduction

Evaluation of information sources (or ‘critical research assessment’ or ‘source criticism’) is becoming a focus in library and information science (LIS), because users have easy access to overwhelming amounts of documents. From being relatively scarce, ‘information’ is becoming abundant. In schools of LIS, courses in ‘information competency’ and ‘source criticism’ are becoming common, and much research on evaluating Wikipedia and other kinds of information source is also being carried out, as well as more general research about conceptualizations and methodologies in these fields.

The focus of evaluation in this context is on whether or not a given source is reliable for use in a scholarly argument. Information specialists should consider themselves as a kind of consumer adviser, who provides guidance to users of information sources (the users being mainly students). In this connection it is important to consider ‘the ‘functional view of information sources’, which states that a source is not in itself good or bad, but just more or less fruitful or relevant in relation to a given question.

Research and practice related to the evaluation of information sources are not independent of general approaches or ‘paradigms’ in LIS. In other articles I have argued that LIS should be based on a social, epistemological approach rather than on an individual, cognitive view. The present article is first of all inspired by Marc Meola’s criticism [1] that the popular checklist approach is based on a ‘mechanistic way of evaluating that is at odds with critical thinking’, and his attempt to provide alternative ways of evaluation. Second, the present paper is inspired by Bailin and Grafstein [2], who based their evaluation very much on real-life examples from the scholarly literature, and like myself are much concerned with ‘paradigm theory’.

In the following, 12 different approaches to evaluating information sources are presented and discussed, together with the core literature and criticism of each approach. It is not an in-depth treatment of each approach but – as stated in the subtitle – an annotated catalogue of different approaches meant to serve students as well as teachers and researchers in this field. A similar overview has not been published before, and the different approaches are mostly used without considering each other.

Corresponding author:

Birger Hjørland, Royal School of Library and Information Science, 6 Birketinget, DK-2300 Copenhagen S, Denmark.

Email: bh@iva.dk

2. The checklist approach

In this approach a list of relevant points to be examined is constructed. They often include issues such as authority, accuracy, objectivity, currency, and coverage of a document. The points may or may not be ranked. The list is then applied to a given information source to be evaluated.

Examples from the literature include Beck [3], Cooke [4], and Skov [5] (the last one is in Danish). Skov's list was translated into English in [6]. (See Table 1.)

Criticism. Criticism points out that checklists have a built-in tendency to mix more or less essential criteria, and that there is no necessary relation between the quality of formal attributes (e.g. spelling) that are easy to check, and more central qualities (such as the validity of claims made in the source). If an information source is screened out because of poor spelling and graphics, we may talk of a 'false negative' if the conclusions in the source are valid: the source is deemed bad although it might be fruitful. It is the central qualities that are most important for the user to check, and in this respect the checklist often fails (cf. [6]). Marc Meola writes:

The checklist model rests on faulty assumptions about the nature of information available through the web, mistaken beliefs about student evaluation skills, and an exaggerated sense of librarian expertise in evaluating information. The checklist model is difficult to implement in practice and encourages a mechanistic way of evaluating that is at odds with critical thinking [1, p. 331].

A serious problem with the checklist approach is also that many pseudoscientific web pages (for example those that attempt to disprove evolution, claim that the Holocaust did not take place, or misrepresent historical figures like Martin Luther King on hate sites) often fulfil all the formal requirements put forward in checklists. This means that applying a checklist to such sites often results in a failure to identify false evidence. Here the problem is 'false positive': something is found to be fine although it is wrong.

3. Classical peer review

In this approach, usually two or three subject experts evaluate an information source. The subject experts are supposed to be experts at the same level as the author of the source to be evaluated (hence the term 'peer'). The reviews may be blind

Table 1. Checklist for evaluating a web resource

<ul style="list-style-type: none"> ● Fitness for purpose <ul style="list-style-type: none"> – Does the source contain information about the purpose and target group? – Does the page fulfil its purpose? ● Authority/trustworthiness <ul style="list-style-type: none"> – Is it possible to identify the author/originator? – What are the qualifications of the originator? – What is the publishing institution/organization? – Is there a contact address? – Does the source contain opinions or facts? – Does the source contain special pleading? – Is there advertising? Is any sponsor named? – Are there references to sources or to literature? – How is the quality control maintained? ● Content <ul style="list-style-type: none"> – Does the source contain unique information? – What is the coverage like (depth and comprehensiveness)? – Is the information relevant? Are there annotated links to other pages? – Is the information correct? – Are there misspellings and signs of botched work? – Does the source contain unfinished sections? 	<ul style="list-style-type: none"> ● Updating <ul style="list-style-type: none"> – Are updating and maintenance done regularly? – Are dates of construction and updating provided? – Does the source contain a news section? – Does the source contain dead links? ● Navigation <ul style="list-style-type: none"> – Are there clear logical structures and good order? – Is it easy to navigate between sections? – Is there a table of contents or a site map? – Are there buttons for up, down, and home? – Is there a standardized, recognizable design? – Is there an internal search engine? ● Design/style <ul style="list-style-type: none"> – Are there relevant graphics? – What are the readability and the legibility of the text like? Is the text worth reading? – Is it of sufficiently good quality to create enthusiasm? ● Availability <ul style="list-style-type: none"> – Is the server busy? – Is there toll access or free access? – Does the source require registration? Does it use cookies? ● Performance <ul style="list-style-type: none"> – Are there heavy graphics? – Is 'text-only' a possibility? – Is the site suitable for the disabled?
---	--

Source: from [6], p. 1893.

or double blind (as opposed to so-called ‘open reviews’). In blind reviews the author does not know who the reviewers are. In double-blind reviews the reviewers are not informed about the identity of the author. In both cases a formal evaluation form may or may not be used. The evaluations are not published.

Peer review is usually regarded as a ‘gold standard’ in academia. In some bibliographic databases it is possible to limit the search to peer-reviewed journals only. (So far, there are no investigations of whether non-peer reviewed journals, such as *Library Trends*, have lower citation rates than peer-reviewed journals; it should be expected that a complex set of different factors are interacting.)

Criticism. Peer review is heavily discussed and criticized, and experiments such as [7–12] indicate that the reliability of the evaluations is low. Chubin and Hackett [13] found in a survey of members of the Scientific Research Society that only 8% agreed that ‘peer review works well as it is’ (p. 192). These criticisms are indeed serious, but the problem is that nobody has yet been able to suggest a better evaluation process that has won broad consent. Other important criticisms are that peer review has a built-in conservatism, and tends to suppress innovative ideas ([14]), and that ‘peers’ are not just peers, but are often competing individuals in the same narrow field; there are therefore also social-psychological issues involved. An empirical study by Mahoney [15] found that reviewers are strongly biased against manuscripts that report results contrary to their theoretical perspective. Also, reviewers are influenced by criteria connected to a certain approach or ‘paradigm’, and may be indisposed to see the quality of papers written from other perspectives. Studies have also found that reviewers often do not identify factual errors in the manuscripts they review.

There is a comprehensive body of scholarly writings about peer reviews. In addition to the literature mentioned above, the following deserve to be considered: Weller [16] is a monograph written in the context of information science; Shatz [17] is a critical inquiry with emphasis on, among other issues, ethical aspects of the review process; Hamermesh [18] examined who the referees are, how they are appointed, whether the assignment is done fairly, and how much time referees spend on the reviews; [19] is a special issue of the *JAMA (Journal of the American Medical Association)* publishing papers from the First International Congress on Peer Review in Biomedical Publication.

4. Modified peer review

Here we are talking about modifications of the classical peer review. The journal *Nature*, for example, used its referees to compare two anonymous sources each (one from Wikipedia and one from *Encyclopedia Britannica*). The reviews were published [20], and this study is today the most cited one about Wikipedia.

Another alternative is to create other groups to carry out the reviews. Miller et al. [21], for example, asked librarians to evaluate Wikipedia.

Criticism. Basically this method has the same inherent problems as classical peer review. *Encyclopedia Britannica* [22] made a criticism of [20], but *Nature* answered this criticism [23] (see also [24]) and *Britannica* did not present new important principal arguments against the peer review method as such.

Hjørland [6] criticized the method used by *Nature* [20], arguing that many scientific issues are controversial; the position of the reviewers is therefore important for the outcome of the review. He therefore suggested an alternative method: examination of how a given source treats controversial views (see below).

5. Evaluation based on examining the coverage of controversial views

A controversy in a given field is identified, and the most important references or spokespersons from each side of the controversy are also identified. How this controversy is covered in the information source to be evaluated is then checked. Does the source mention the controversy? Does it provide references to each side? Examples of the application of this method are [6] and [25].

Criticism. There are some forms of ‘established knowledge’ that are not controversial (e.g. the melting point of lead); in such cases no controversies can be found, but it is still important to check for error. There is also a principal problem in determining when a point of view is serious enough to be taken into consideration. Finally, the usefulness of the method depends on the identification of controversies, which presupposes good knowledge of the subject literature in the field. It is time consuming, but the time might be well spent.

6. Evidence-based evaluation

This type of evaluation is based on the principles of evidence-based practice (EBP), which is an interdisciplinary movement that started as evidence-based medicine (EBM) in about 1992. Sackett et al. [26] give the following definition: ‘Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients.’ Such a definition is, however, impossible to differentiate from other forms of scientifically based medicine. It has therefore been claimed [27] that EBM – as well as evidence-based practice in general – is based on certain ideas of what counts as the best evidence, and here the movement is generally interpreted as a movement with a strong affiliation to positivist ideals of science competing with other views about what counts as the best evidence [27]. The most characteristic ideal in EBP is the administration of a hierarchy of research methods that is independent of the questions being answered. It is basically characterized by the neglect of theoretical and qualitative research.

Applied to the evaluation of information sources, this means that evidence-based evaluation examines the evidence on which the conclusions of a given source are based. If it is a primary source (e.g. a research article), emphasis is put on evaluation of the research method applied (mostly from a version of the evidence hierarchy). If it is a secondary or tertiary information source (e.g. a systematic review or a Wikipedia article), it is evaluated by the principles of meta-reviews – that is, from an examination of its sources and the principles of synthesizing research results.

This approach is essentially based on the use of research methodology to evaluate documents. Textbooks on research methodology are written in order to train students to do research. They may, however, just as well be used in order to evaluate research. Some books are explicitly written for this purpose. For example:

- *Studying a study and testing a test: How to read the medical evidence* [28];
- *Evaluating research articles: From start to finish* [29];
- *How to read a paper: The basics of evidence-based medicine* [30];
- *Evaluating information: A guide for users of social science research* [31];
- *Questioning numbers: How to read and critique research* [32]; and
- *Reasoning with statistics: How to read quantitative research* [33].

Whereas many such books deal with narrow empirical methods, some include also theoretical, ideological, and critical aspects:

- *What's behind the research? Discovering hidden assumptions in the behavioral sciences* [34];
- *Sociological theory – A contemporary view: How to read, criticize and do theory* [35]; and
- *How to Read Karl Marx* [36].

This last example makes it obvious that any prescription of how to read depends on the theoretical stance taken by the author of that prescription (which is also demonstrated by the overwhelming number of guides on how to read the Bible). In the end, the criteria for evaluation of research are a matter of theories in the philosophy of science. Critical debates on these issues are taking place in the epistemological discourses.

Criticism. Evidence-based evaluation is criticized as being too narrow, too mechanical, and too formalist [27]. It is acknowledged that practice should always be based on the best evidence, and on relevant scientific knowledge. Non-positivist epistemologies such as some versions of pragmatism, hermeneutics and critical theory provide a broader and less mechanical way to evaluate information sources – not just in the humanities and social sciences, but also in sciences such as medicine. When such broader epistemological questions are taken into account, we should speak of research-based evaluation (rather than evidence-based evaluation), which is the core of scholarly criticism.

7. Comparative studies

The coverage of a given subject in a document may be compared with the coverage of the same subject in a small number of ‘authoritative works’ in the field. Based on these works, a comprehensive table of content of subtopics and a scheme in which the table of contents appears for each of the works may be made (see examples in [37] comparing information about major philosophers). For each of the sources a checkmark is made in order to indicate whether the subtopic is covered in that work.

Subtopics that are covered by all the authoritative sources are considered mandatory, and the work under examination is evaluated by counting the percentage of mandatory subtopics it includes. In the example in Table 2, subtopics 2, 3

Table 2. Scheme for comparing coverage in different information sources

	Work 1	Work 2	Work 3	Work 4 (to be examined)
Subtopic 1		×	×	×
Subtopic 2	×	×	×	
Subtopic 3	×	×	×	×
...		×		
Subtopic N	×			

and 4 are covered by all the authoritative sources, and the work to be examined covers one of these subtopics, or 33%. Alternatively – or additionally – the sources may be examined by counting the numbers of errors they contain (see for example [38]).

Meola [1] also contains important arguments and examples on how to evaluate information sources by using other sources for comparison and fact-checking.

Criticism. There is a problem in determining which works to recognize as being ‘authoritative’ (and thus suitable for comparison), which demands a good working knowledge of the sources in the field. The fact that works are seldom independent, but are used by authors of new works, should also not be underestimated. The method is most suitable in research fields that are not too dynamic (after a scientific breakthrough the concept of ‘authoritative work’ is problematic).

8. Author credentials

Bailin and Grafstein [2] consider the study of author credentials to be one of three gold standards in research assessment (the other two are peer review and publisher reputation). It is not enough for an author to have a high academic degree (e.g. a PhD) or affiliations with a university; rather the author’s credentials should normally be in the specific field of the document to be evaluated. Even receivers of the highest academic recognition (e.g. Nobel Prize winners) may speak or write about matters outside their competence.

Ways of examining author credentials include:

1. looking up the author’s CV on his or her home page;
2. considering the biographies often provided in books or in conference presentations; and
3. considering his or her bibliometric data (publications as well as citations).

The criteria for author credentials are by principle the same as the criteria used in selecting reviewers for peer review: the author should be *active* in the research field in question, possibly a senior researcher, and should have no conflicts of interest. Even when these criteria are fulfilled, there still remains the question of the author’s theoretical orientation.

Sometimes different factors, such as sex, ethnicity or religion, have been considered. Some feminist researchers may, for example, consider it important that the author be a woman. In evaluating single information sources this is generally a very bad idea, because the best publications may be deselected (i.e., ‘false negatives’). However, in research policy it is relevant to consider whether there is a broad representation of different views among researchers, including a balanced representation of sexes, ages, ethnic groups, and so on. This is obviously more important in the case of, for example, history compared with chemistry.

Criticism. First, to evaluate a document by considering the credentials of its author is only an indirect evaluation. After all, it is the properties of the document itself that matter. Qualified authors sometimes write bad papers, and the converse. Second, new scientific breakthroughs are often made by ‘outsiders’ (see, for example, [2], [39] and [40]). Third, researchers in a field may also attempt to make the impression that consensus is stronger than it is; they may have an interest in not questioning their own field as much as is necessary – depending on the culture in that field.

Bailin and Grafstein [2] present and discuss the case of hormone replacement therapy (HRT) for women. The research first suggested that HRT was beneficial, but later research suggested that, in addition to the benefits, HRT poses serious dangers. Nevertheless, the research was published by researchers with appropriate credentials.

9. Publisher reputation

Bailin and Grafstein [2] also consider publisher reputation as one of the three gold standards in research assessment. Journals and monographs as well as 'grey literature' may be evaluated according to the status of their publishers. In the case of journals three aspects may be considered:

1. Is the journal published by a publisher with a fine reputation?
2. Is the journal itself well respected?
3. What is the journal's impact factor (JIF)?

Question 2 is more important than question 1: if we know that a journal is, for example, considered a leading one in its field, we need not ask more. The question of JIF is considered separately below. Two issues in particular are important to consider. First, journals are often considered to form a prestige hierarchy: some journals are top quality, some are low quality, and some are in the middle. Papers not accepted by top journals are usually accepted elsewhere. Second, it is a mistake to believe that all journals in an academic discipline form a single hierarchy; they are often specialized and do not compete for the same articles as journals representing other subfields. Also, as shown empirically by Heine Andersen [41], researchers have different perceptions of those hierarchies – that is, of what constitutes the most important journals. 'Paradigms' again play an important role in this respect.

In relation to the evaluation of monographs (and anthologies), publishers play a more direct role. Vanity presses are often considered to be a sign of bad quality, while highly respected publishers are considered to be an indicator of good quality. University presses are generally considered to be a sign of quality, and they have mostly adopted a peer review process, because they know that their authors are mostly under pressure to publish in peer-reviewed sources.

It is possible to rank publishers by bibliometric indicators, but this is less important in disciplines that are dominated by journal publications (in which case the JIF may be used as a better alternative). Concerning grey literature, bibliometric data indicating the status of the departments issuing the documents are a hypothetical possibility.

Publishers may have special economic, ideological or disciplinary interests or profiles. This aspect is discussed in the section about sponsoring below.

Criticism. As with author credentials, evaluation of a document by evaluating its publisher is an indirect kind of evaluation. The same publisher may publish individual documents with varying degrees of quality: that a document is published by a publisher with a fine reputation is never a guarantee of quality. As already stated, Bailin and Grafstein [2] present and discuss the case of HRT for women. The research first suggested that HRT was beneficial, but later research suggested that, in addition to the benefits, HRT poses serious dangers. Nevertheless, the research was published in very respectable publications.

10. Journal impact factor

Journal impact factor is one kind of impact factor (IF); another kind is web impact factor (WIF). JIF is a statistical measure of trenchancy, measured by citations received in other journals. It is calculated each year by Thomson Reuters (formerly the Institute for Scientific Information, ISI) and is termed the Thomson Reuters Impact Factor and published in its *Journal Citation Reports* [42]. The JIF of a given journal is calculated by dividing the number of citations received in the current year by the number of source items published in that journal during the previous two years. The JIF is a tool for ranking *journals*, but it is also marketed as a tool for evaluating single *articles*:

Perhaps the most important and recent use of impact is in the process of academic evaluation. The impact factor can be used to provide a gross approximation of the prestige of journals in which individuals have been published. This is best done in conjunction with other considerations such as peer review, productivity, and subject specialty citation rates. [43]

And:

Impact factor is not a perfect tool to measure the quality of articles but there is nothing better and it has the advantage of already being in existence and is, therefore, a good technique for scientific evaluation. Experience has shown that in each specialty the best journals are those in which it is most difficult to have an article accepted, and these are the journals that have a high impact factor. Most of these journals existed long before the impact factor was devised. The use of impact factor as a measure of quality is widespread because it fits well with the opinion we have in each field of the best journals in our specialty. [44], cited in [45]

There is therefore a tendency to use JIF as an indicator of journal quality, and also as an indicator of the quality of each article published in a given journal.

Criticism. There is a comprehensive literature criticizing the JIF – especially its use as an indicator of the quality of articles and/or their authors. The Norwegian cancer researcher Per Ottar Seglen is one of the main critics of this measure. Research by his team was pioneering in the field of autophagy, and thanks to their contributions the Norwegian Radium Hospital has the opportunity to stay at the forefront of this research field. This was not, however, the picture shown by bibliometric indicators. Therefore Professor Seglen decided to examine the nature of and mechanisms behind the bibliometric indicators, and he became one of the leading critical bibliometricians, in particular in relation to the JIF. Among his papers in this field are [46], [47] and [48]. In his article ‘Why the impact factor of journals should not be used for evaluating research’, Seglen concludes:

Summary points:

- Use of journal impact factors conceals the difference in article citation rates (articles in the most cited half of articles in a journal are cited 10 times as often as the least cited half)
- Journals’ impact factors are determined by technicalities unrelated to the scientific quality of their articles
- Journal impact factors depend on the research field: high impact factors are likely in journals covering large areas of basic research with a rapidly expanding but short lived literature that use many references per article
- Article citation rates determine the journal impact factor, not vice versa [48].

Another medical researcher conducted a review of studies on JIF and concluded:

In sum, the impact factor is ill conceived, intellectually and technically flawed, and [a] misleading effort to assess the academic worth of a paper [Walter et al., 2003] [49]. And, if that’s not enough, Walter and colleagues (2003) point out that the impact factor (IF) has now ‘... spawned a range of flawed offspring, including “Scope-adjusted IF”, “Discipline-specific IF”, “Journal-specific influence factor”, “Immediacy index” and “Cited half-life”’. Exercise physiologists should reflect critically on the issues presented in this article. This article is by no means complete in its analysis, it is nonetheless reasonably complete to help with the decision to make a clean break from the use of impact factors. [50]

A historical review of the development of the JIF is ‘Garfield and the impact factor’ [51]. The debates about JIF continue, and a recent contribution from the field of organizational studies is [52].

II. Sponsoring: tracing the influence of economic, political and ideological interests

Bailin and Grafstein [2, p. 38] write: ‘Research is generally dependent on financial support, what gets funded to some degree determines what gets researched and thus what research questions are asked. Underlying any research question are certain assumptions about the nature of the phenomena being examined.’ The book provides three examples of how funding has influenced research:

- Case 1: hormone replacement therapy (HRT);
- Case 2: Enron;
- Case 3: *The bell curve*.

In all three cases the book argues that the sponsoring behind the research has hindered the establishment of the truth. In the HRT case the pharmaceutical industry supported research that suppressed findings of serious side effects. In the Enron case the auditing firm (Arthur Andersen) had a conflict of interest, because it was dependent on major income from Enron (of which the smallest part was for the audit itself). Therefore nobody at Arthur Andersen ‘blew the whistle’ when Enron carried out fraudulent trading. In the case of *The bell curve* it was shown that the authors (Richard J Herrnstein and Charles Murray) received financial support from the American Enterprise Institute, which is conservative, and has an ideological interest in favour of the book’s findings. ‘The research findings in *The bell curve*’, Bailin and Grafstein [2] write, ‘were not based on disinterested scientific inquiry, but ... on the contrary, the book was written with a clear political agenda in mind’ (p. 36).

Criticism. As with author credentials and publisher reputation, to evaluate a document by tracing its sponsorships is an *indirect kind of evaluation*. If it should have any implications it should be made probable that the sponsorship has influenced the conclusions in a negative way. The evaluation of a work by examining its dependence on sponsors, although it is an indirect method, is, however, an important part of the critical work. Although there are no necessary relations between sponsoring and conclusions, conflicts of interest should always be carefully examined.

12. Book reviews

Book reviews can, depending on their quality, be used to evaluate books as information sources. Also, the method used to write good book reviews may serve as a valuable evaluation method (cf., [17, pp. 109–120]). Book reviews have been termed ‘published peer reviews’ [53], and George Sarton [54], a historian of science, summarized the qualities of a good book review: a review should describe and characterize not only the book in question, but also the subject with which it is dealing. It should furthermore answer the question ‘Is the book a real addition to our knowledge, and if so, what exactly has been added?’

The Danish historian Svend Ellehøj [55] wrote about the criteria that are relevant in the review of historical monographs. The review should not just consider the research problem and structure of the argumentation, and its relation to the relevant primary sources and literature in the field; as far as possible it should also inform about the author, his or her environment, and his or her relation to more general theories. Ellehøj also emphasizes that, although the set of criteria that are active in book reviews is extremely complicated, it is nonetheless possible to produce evaluations of historical research results that have a satisfactory level of validity.

Good book reviews (although not the average book review) may cause the reader to completely change his or her opinion of a specific book. A good review informs about the strong and weak aspects of a book compared with other books on the same subject. It evaluates the sources used: are they the most important ones available? It also evaluates the interpretation of the sources, the scholarly perspective, and the validity of the conclusions. In the best book reviews, the evaluation of the details and of the metatheoretical perspective may be mutually supportive.

Criticism. It is hard to find any objections in principle to the book review as a method of information evaluation. It is easy, however, to criticize the level of most book reviews; they have even been branded a second-class citizen of scientific literature [56]. Book reviews are, moreover, regularly charged with merely reflecting individual opinions, which, according to their critics, disqualifies them entirely as scholarly contributions [57]. However, none of this criticism detracts from the conclusion that, when done properly and qualified, book reviews and the methodology of book reviewing are a first-class scholarly tool.

13. Broader criteria

A single document to be evaluated is always part of broader contexts. It needs to be seen in a historical and social perspective. The understanding of the single document should (according to the hermeneutical circle) be supplemented by understanding of the culture or domain in which the document is produced (and vice versa).

Again, the example in which Bailin and Grafstein [2] present and discuss the case of hormone replacement therapy (HRT) for women should be mentioned. In spite of its problematic findings, all of the research met the gold standards: it was peer-reviewed, the research was published in very respectable publications, and the researchers had appropriate credentials. This indicates that the criteria used are influenced by background knowledge and theory, and when this changes then former evaluations are no longer valid. This is an argument for the hermeneutic philosophy of science, and against the positivist philosophy of science.

In the humanities, *reception studies* are the study of the influence of a given work or a given author (e.g. the influence and reception of Friedrich Nietzsche). Reception studies work from the premise that all texts are historically situated, and acquire new meanings in new contexts; they demonstrate that the influence of a work is dependent not just on the internal properties of that work but also on the different cultures, social groups and disciplines in which it has been used, ignored, or rejected. It is often the case that something that has formerly been refused may be re-evaluated and receiving new attention. This is for example the case with the field of classical rhetoric, which was broadly rejected in the age of logical positivism, but which since then has again become a highly respected and influential field. The lesson for the present article is again that individuals evaluating sources always do so from a perspective or ‘paradigm’, and that it is not enough to focus on narrow criteria for evaluating evaluation sources.

14. Conclusion

After receiving tuition in source criticism, a student wrote: ‘I now know that to evaluate a text is not “just” to read it, understand it, and to criticize its content. More professional methods are needed.’

This sentence made me a little worried about the consequences of teaching the methods discussed in this article. I do not believe that any of these methods can do the job, alone or in combination. I think the goal is to teach students to read texts, to understand them, and to be able to provide a relevant criticism of them. That is the end goal, and that should be sufficient. All the methods are ‘just’ steps on our way in learning how to read, understand, and criticize texts. But as such they may represent valuable or indeed necessary steps that counteract the use of narrow and one-sided approaches. To read a text is often not a simple process. It involves considering a text in relation to other texts, in relation to the methods used and the social interests it serves and counteracts, and in relation to its public reception as indicated by, for example, bibliometric measures. In other words, it involves the subjectivity of the reader, and that subjectivity is influenced – more or less – by different theoretical perspectives. Good, scholarly reading is to be aware of different perspectives, and to situate oneself among them.

References

- [1] Meola M. Chucking the checklist: a contextual approach to teaching undergraduates web-site evaluation. *Portal: Libraries and the Academy* 2004; 4(3), 331–344.
- [2] Bailin A, Grafstein A. *The critical assessment of research: Traditional and new methods of evaluation*. Oxford, UK: Chandos, 2010.
- [3] Beck SE The good, the bad and the ugly, or, why it’s a good idea to evaluate web sources. Web-page developed at New Mexico State University Library 1997 (Retrieved 10 December 2011 from <http://lib.nmsu.edu/instruction/eval.html>)
- [4] Cooke A. *A guide to finding quality information on the internet: Selection and evaluation strategies*. 2nd ed. London, UK: Library Association, 2001.
- [5] Skov A. Criteria for evaluating web-resource 2000 (translated from the Danish by Birger Hjørland from ‘Kvalitetsvurdering af web-sider’). (Retrieved 10 December 2011 from <http://www.iva.dk/jni/lifeboat/info.asp?subjectid=308>; original Danish version retrieved from http://vip.iva.dk/tutorials/kvalitet/ny_side_2.htm)
- [6] Hjørland B. Evaluation of an information source illustrated by a case study: effect of screening for breast cancer. *Journal of the American Society for Information Science and Technology* 2011; 62(10): 1892–1898.
- [7] Cicchetti DV. The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences* 1991; 14(1): 119–135.
- [8] Cole S, Cole JR, Simon GA. Chance and consensus in peer review. *Science* 1981; 214(4523): 881–886.
- [9] Garfunkel JM, Ulshen MH; Hamrick HJ, Lawson EE. Problems Identified by Secondary Review of Accepted Manuscripts. *JAMA (Journal of the American Medical Association)* 1990; 263(10): 1369–1371.
- [10] Peters DP, Ceci SJ Peer-review practices of psychological journals: the fate of published articles, submitted again. *Behavioral and Brain Sciences* 1982; 5(2): 187–255.
- [11] Rothwell PM, Martyn, CN. Reproducibility of peer review in clinical neuroscience. Is agreement between reviewers any greater than would be expected by chance alone? *Brain* 2000; 123(9): 361–376.
- [12] Starbuck WH. How much better are the most prestigious journals? The statistics of academic publication. *Organization Science* 2005; 16(2): 180–200.
- [13] Chubin M, Hackett, E. *Peerless science: Peer review and US science policy*. New York, NY, USA: State University of New York Press, 1990.
- [14] Horrobin DF. The philosophical basis of peer review and the suppression of innovation. *JAMA (Journal of the American Medical Association)* 1990; 263(10): 1438–1441.
- [15] Mahoney MJ. Publication prejudices: an experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research* 1977; 1(2): 161–175.
- [16] Weller AC. *Editorial peer review: Its strengths and weaknesses*. Medford, NJ, USA: Information Today, 2001 (ASIS&T Monograph Series).
- [17] Shatz D. *Peer review: A critical inquiry*. Lanham, MD, USA: Rowman and Littlefield Publishers, 2004 (Issues in Academic Ethics).
- [18] Hamermesh DS. Facts and myths about refereeing. *Journal of Economic Perspectives* 1994; 8(1): 153–163.
- [19] Rennie D. Editorial peer review in biomedical publication: the first international congress. *JAMA (Journal of the American Medical Association)* 1990; 263(10): 1317 (Thematic issue: ‘Guarding the Guardians’ about research in editorial ‘Peer Review’).
- [20] Giles J. Special report: Internet encyclopaedias go head to head. *Nature* 2005; 438: 900–901 (Retrieved 8 December 2011 from: <http://www2.stat.unibo.it/mazzocchi/macroecnomia/documenti/Nature%202006%20wikipedia.pdf>).
- [21] Miller BX; Helicher K, Berry, T. I want My Wikipedia! *Library Journal* 2006; 131(6): 122–124 (Retrieved 8 December 2011 from: <http://www.libraryjournal.com/article/CA6317246.html>).

- [22] Encyclopedia Britannica. Fatally flawed. Refuting the recent study on encyclopedic accuracy by the journal *Nature*. 2006 (Retrieved 8 December 2011 from http://corporate.britannica.com/britannica_nature_response.pdf).
- [23] Giles J. Supplementary information to accompany *Nature* news article 'Internet encyclopaedias go head to head'. 2005 (Retrieved 8 December 2011 from: <http://www.nature.com/nature/journal/v438/n7070/extref/438900a-s1.doc>).
- [24] Lubick N. Wikipedia v. Britannica: Reader beware. *Geotimes* 2006; 51(4): 46 (Retrieved 8 December 2011 from: <http://www.agiweb.org/geotimes/apr06/geomedia.html#FIRST>).
- [25] Luyt B. The nature of historical representation on Wikipedia: dominant or alternative historiography? *Journal of the American Society for Information Science and Technology* 2011; 62(6): 1058–1065.
- [26] Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *British Medical Journal (BMJ)*; 1996; 312(7023): 71–72 (Retrieved 8 December 2011 from: <http://www.bmj.com/content/312/7023/71.full>).
- [27] Hjørland B. Evidence-based practice: an analysis based on the philosophy of science. *Journal of the American Society for Information Science and Technology* 2011; 62(7): 1301–1310.
- [28] Riegelman RK. *Studying a study and testing a test: How to read the medical evidence*. 5th ed. Philadelphia, PA, USA: Lippincott Williams and Wilkins, 2004.
- [29] Girden ER. *Evaluating research articles: From start to finish*. London, UK: SAGE Publications, 1996.
- [30] Greenhalgh T. *How to read a paper: The basics of evidence-based medicine*. 4th ed. Oxford, UK: BMJ Books, 2010.
- [31] Katzner J, Cook KH, Crouch WW. *Evaluating information: A guide for users of social science research*. 4th ed. Boston, MA, USA: McGraw-Hill, 1998.
- [32] Wilkins KG *Questioning numbers: How to read and critique research*. New York, NY, USA: Oxford University Press, 2010.
- [33] Williams F, Monge PR. *Reasoning with statistics: How to read quantitative research*. 5th ed. Florence, KY, USA: Wadsworth Publishing Company, 2000.
- [34] Slife BD, Williams RN. *What's behind the research? Discovering hidden assumptions in the behavioral sciences*. London, UK: SAGE Publications, 1995.
- [35] Smelser NJ. *Sociological theory – a contemporary view: How to read, criticize and do theory*. New Orleans, LA, USA: Quid Pro Books, 2011 (1st ed. 1971).
- [36] Fischer E, Marek F. *How to read Karl Marx*. New York, NY, USA: Monthly Review Press, 1996.
- [37] Bragues G. Wiki-philosophizing in a marketplace of ideas: evaluating Wikipedia's entries on seven great minds. *MediaTropes eJournal* 2009; 2(1): 117–158.
- [38] Rector LH. Comparison of Wikipedia and other encyclopedia for accuracy, breadth, and depth in historical articles. *Reference Services Review* 2008; 36(1): 7–22.
- [39] Fleck L. *Genesis and development of a scientific fact*. Chicago, IL, USA: University of Chicago Press, 1979 (German original 1935).
- [40] Hjørland B, Nicolaisen J. The social psychology of information use: seeking 'friends', avoiding 'enemies'. *Information Research* 2010; 15(3) colis706 (Available at <http://InformationR.net/ir/15-3/colis7/colis706.html>).
- [41] Andersen H. Influence and reputation in the social sciences: how much do researchers agree? *Journal of Documentation* 2000; 56(6): 674–692.
- [42] *Journal Citation Reports*. Philadelphia, PA, USA: Thomson Reuters (annual). Official website: http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports/
- [43] Garfield E. The impact factor. Current Contents print editions June 20, 1994. Here cited 8 December 2011 from: http://thomsonreuters.com/products_services/science/free/essays/impact_factor/
- [44] Hoeffel C. Journal impact factors. *Allergy* 1998; 53(12): 1225.
- [45] Garfield E. The agony and the ecstasy: the history and meaning of the journal impact factor. Presented at the *5th International Congress on Peer Review and Biomedical Publication*, Chicago, IL, USA, 16 September 2005 (Retrieved 9 December 2011 from: <http://www.garfield.library.upenn.edu/papers/jifchicago2005.pdf>).
- [46] Seglen PO. How representative is the journal impact factor? *Research Evaluation* 1993; 2(3): 143–149.
- [47] Seglen PO. Bruk av siteringer og tidsskriftimpaktfaktor til forskningsevaluering. *Biblioteksarbejde* 1996; 17(48): 27–34 (Retrieved 9 December 2011 from: [http://biblioteksarbejde.dk/art/BA48/seglen\(1996\).pdf](http://biblioteksarbejde.dk/art/BA48/seglen(1996).pdf)).
- [48] Seglen PO. Why the impact factor of journals should not be used for evaluating research. *British Medical Journal* 1997; 314(7079): 498–502 (Retrieved 10 December 2011 from: <http://www.bmj.com/content/314/7079/497.1?tab=full>).
- [49] Walter G, Bloch S, Hunt G, Fisher K. Counting on citations: a flawed way to measure quality. *Medical Journal of Australia* 2003; 178(6): 280–281.
- [50] Boone T. Journal impact factor: a critical review. *Professionalization of Exercise Physiology* [online journal] 2004; 7(1) (Retrieved 10 December 2011 from: <http://faculty.css.edu/tboone2/asep/journalIMPACTfactor.html>).
- [51] Bensman SJ. Garfield and the impact factor. *Annual Review of Information Science and Technology* 2007; 41: 93–155.
- [52] Baum JAC. Free-riding on power laws: questioning the validity of the impact factor as a measure of research quality in organization studies. *Organization* 2011; 18(4): 449–466.
- [53] Hyland K. *Disciplinary discourses: Social interactions in academic writing*. Harlow, UK: Longman, 2000.
- [54] Sarton G. Notes on the reviewing of learned books. *Science* 1960; 131(22 April): 1182–1187.

- [55] Ellehøj S. Faglige kriterier for vurdering af historiske forskningsresultater i: Problemvurdering og prioritering i historie. *Studier i historisk metode V* 1970: 72–80 (in Danish).
- [56] Riley LE, Spreitzer EA. Book reviewing in the social sciences. *The American Sociologist* 1970; 5(November): 358–363.
- [57] Sabosik PE. Scholarly reviewing and the role of choice in the postpublication review process. *Book Research Quarterly* 1988; 4(2): 10–18.